Sentiment Analysis in E-Commerce: A Review on The Techniques and Algorithms

Muhammad Marong
School of Computing and Technology
Asia Pacific University of Technology & Innovation
Malaysia
mahamarong@gmail.com

Nowshath K Batcha
School of Computing and Technology
Asia Pacific University of Technology & Innovation
Malaysia
nowshath.kb@apu.edu.my

Raheem Mafas
School of Computing and Technology
Asia Pacific University of Technology & Innovation
Malaysia
raheem@staffemail.apu.edu.my

***Abstract*—The internet and social media platforms have made available massive quantities of information to users worldwide. Numerous internet sources are present on categorical views of activities, goods and services, beliefs or perhaps the mood created by the online dwellers. In this competitive business world, various industries especially e-commerce immensely use sentiment analysis to increase productivity and make better business decisions. Sentiment Analysis is an associate degree in the field of analytics which has proven to be one of the significant instruments to reveal actionable insights using very big text databases from plentiful domains. This paper tackles a comprehensive overview of sentiment analysis and relevant techniques in e-commerce sector that is always keen to find out about the consumers' opinions of their goods and services. It starts with the notion of an assessment of sentiments that have emerged as a method for understanding clients' emotions. It also describes the normal assignment engaged in methods of sentiment analysis, such as methods based on lexicon and monitored machine learning. However, sentiment analysis is applied within the retail business largely in the E-Commerce that allows the operators to enhance their business operation.**

Keywords— E-Commerce, sentiment analysis, lexical approach, machine learning, data mining.

I. INTRODUCTION

The expression on major news announcements can have a high impact on the financial market and investor behavior resulting in rapid changes or abnormal effects in financial portfolios. In the present world, the availability and exponential increase in the use of the Internet have resulted in individuals preferring to communicate and share data on various subjects ranging from end-products to various services including healthcare. The basic idea behind these new emerging analytics technologies is to understand, predict human behavior & attitude and provide information to the traders that could be used to foresee and organize the business process before making an investment or risk management decisions.

E-Commerce destinations are the market capital. Because of the reliability of locations, no one needs to go outside the industry, where these locations are more reliable to the point of being showcased [1]. Customers look into e-commerce locations as they sit tight for the products to go ahead. At this point, a big amount of settled and rumored organizations' websites are propelling their products, maintaining trust in these locations is the ultimate objective, and Sentiment Analysis is a main concern in this region. The term sentiment meaning a view or opinion expressed and analysis meaning structure of something, hence putting these two words in one meaning helps to uncover those feelings.

Sentiment Analysis is a form of Natural Language Processing (NLP) which tracks the mood and attitude of the public regarding any item or topic [2]. Sentiment Analysis or Opinion Mining, which includes the construction of a scheme or model for identifying and studying data aimed at obtaining and examining people's feelings expressed positively or negatively through the analysis of a big quantity of data from surveys, reactions and reviews. It is a field of study that could be useful in many ways. For example, in marketing, it helps to provide better product analytics or even monitor market research that can determine which version of a product or service are problematic or popular.

II. SENTIMENT ANALYSIS IN E-COMMERCE

Today, the rapid development of Internet and its users has altered the way how individuals communicate worldwide, in particular when doing business, the internet applications on business operations have developed new possibilities how products or services are sold in the globe today. Accessibility of social media platforms empowered the internet users to express and share their opinions on different kinds of components based on their life experience, including products and services that they enjoy.

Sentiment Analysis has been a burgeoning technology that taps into customer demands based on Natural Language Processing. This motivation is usually used to properly understand what customers want, when, why and how they want it, retailers need to pivot toward sentiment analysis, hence avoid doing the same mistakes and choosing the right decisions based on comments or reviews. As part of e-commerce, online shopping is a good example of how products or services are sold over the Internet. Big name distributors like Amazon and Alibaba along with tiny distributors out there certainly had disappointing outcomes, one of the primary factors for their slow sales was poor product assortment. These distributors were essentially unable to put the correct products on the shelf, and

customers punished them by spending their cash elsewhere [3] as it happened a tech firm like Fitbit in the year 2016 [4].

Consumer understanding has always been high on the to-do list of distributors and the use of sentiment analysis to monitor those emotions was the main motive for businesses to understand how diverse and thorough the opinion mining on the clients' reviews can be. The Internet is a minefield of perspective, being able to access these opinions on a variety of different platforms is a significant advantage for any business seeking to improve its products or services.

III. Sentiment Analysis: Techniques

There are many applications and enhancement on Sentiment Analysis algorithms that were proposed and used from several years to date. This paper aims to give a closer look at the most common techniques used in retail business especially in the E-Commerce sector and give a comprehensive review about it. In sentiment classification, there are two main study fields such as Machine Learning and Lexicon, and each field has its own subdivision as shown in Fig.1. There have also been few studies combining these two techniques and gaining comparatively better efficiency in the sentiment analysis operation [5].

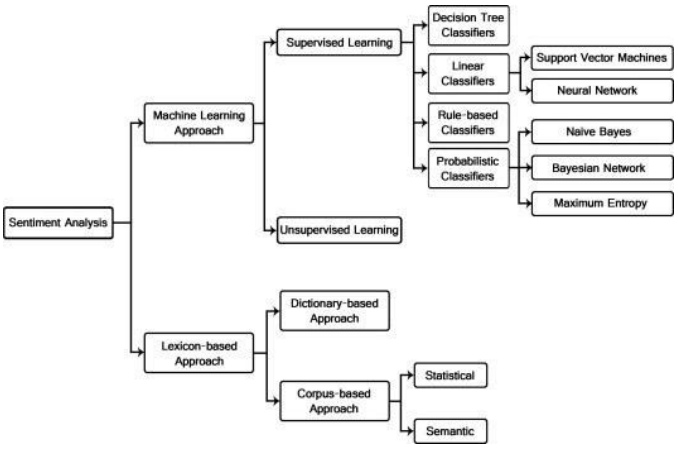

Fig. 1.Sentiment classification techniques [6]

A. Machine Learning approach

Various techniques could be used to do sentiment analysis among which one of the techniques which are commonly used in the retail industry is the Naïve Bayes Method. It is an algorithm of probabilistic learning derived from the principle of Bayesian choice Naïve Bayes Classifier (NBC) would merge fresh understanding with prior understanding. These classification algorithms are simple and have similar effects with other techniques. In the NBC, probability of a message x in class c, is calculated using the following equation [7].

Formula:

$$P(c|x) = \frac{P(x|c)\,P(c)}{P(x)}$$

- P(c) - the predictor (attribute) likelihood provided by the class (target).
- P(c) - likelihood of the previous class.
- P(x) - the probability of the class provided by the predictor.
- P(x) - previous likelihood of the predictor.

The method in which a review can be classified as a positive one (thumbs up) or a negative one (thumbs down) that are extracted from the sentences and are used for sentiment classification. The algorithm has efficiency in text classification with an accuracy of 83%. Naive Bayes model for very big information sets is simple to construct and especially helpful. Naïve Bayes is known to perform even extremely advanced classification techniques in addition to simplicity [8].

B. Lexicon based approach

The Lexicon-based method utilizes predefined word phrases and opinion idioms where every phrase and idiom is assessed as either positive sentiment or negative sentiment. Most researchers used automatic approaches such as dictionary and corpus to assign the opinion words, but they still manually assign words and sentences in statements of view to guarantee the right assignment of words and sentences. [9]. This rule gives comparative slant esteems to semantically close words, in the case of retail industries like e-commerce sites. There are a lot of uncovering comments and reviews which include slangs and misspelling due to different languages. This situation results in a difficult job for automatic system design and development. In addition, to evaluate the feeling of the remark, prior understanding is needed to classify the opinion's polarity. Two approaches could be used under the lexicon classifier such as Dictionary-based method and Corpus-based method to collect online dictionary with number of opinion statements for respective synonyms and antonyms. New phrases are added to the list of seeds and the technique continues to add the phrases iteratively till no new phrases are found. It was emphasized to use manual check to clean the list lastly [10].

IV. Sentiment Analysis: New Opportunities

The rise of social media against retailers face a vibrant and competitive atmosphere, with retailers seeking greater globalization and competitiveness on blogs and social media platforms that fueled interest in the evaluation of feelings [11]. It is a strong technology with excellent potential to assist organizations to concentrate on significant information in their data that needs a correct mechanism to transform it into understanding.

Several studies show that the percentage of individuals and businesses using social media platform applications as a Customer Relationship Management (CRM) tool has amplified significantly. A lot of reviews, criticisms and commendations are normally published and reported only minutes after the release of a new product. Analysis of these information allows companies to adapt to this growing trend in achieving certain business values such as increasing customer base, customer trustworthiness (loyalty), customer satisfaction (pleasure/happiness) and customer reputation while achieving a higher revenue and profit [12]. On the other hand, by analyzing the strengths and weaknesses of the divergent characteristics of items, as well as discovering the satisfaction rates of other consumers of different products, businesses can use this information as testimonials too.

Customers enjoy being heard, using sentiment analysis to define this view or classify it could have a lot of effects. If we look at E-Commerce sites where the greatest reviews or remarks on all types of products & services, constructing a system for examining those reviews could assist to make better choices and monitor brands, improving customer support, keeping an eye on their competitive or even manage a crisis better if it is likely to happen. There's nothing worse than finding a catastrophe or issue before it's too late. Well, sentiment analysis has helped to uncover all these problems, stress and loss of money or resources.

V. SENTIMENT ANALYSIS: CHALLENGES

In the assessment of raw information such as Word Sense Disambiguation, Negotiations, Comparison, Intensity, and Sarcasm, sentiment analysis models face several difficulties because of the human language nature. First, word sense disambiguation is a model which recognizes a word as positive, however it can reflect a negative sense in other cases. For example, “A small number of students is beneficial for students who prefer studying in a small environment”, But if the reader wishes a larger box, tiny box size can be negative. The model must, therefore, be in a particular domain in order to prevent disambiguation [13] [17].

Comparisons can also trigger confusion, apart from disambiguation. Fossil watch, for instance, it's better than watching diesel. While a good opinion means the word "better", if the paper discusses Diesel watch, it should be regarded as a negative polarity. Word intensity can also exaggerate view that can lead to the assignment of a document in an incorrect class. Negation can make a mistake because it can alter a sentence's polarity from positive to negative. For instance, "there is a good possibility that other religions can dominate a country". This phrase implies more variety that is usually deemed positive, but it can be negative because the author is concerned about a good opportunity. Sarcasm is difficult because it needs thorough study because sarcasm uses positive words, but the author implies adverse context [14] [18]. In another scenario, an opinion term that is deemed positive can be regarded as negative because individuals do not express their views in the same way. Some reviews may be simple to comprehend for humans but hard to fix owing to context for computers.

Sentiment Analysis is still not commonly used as it depends on enormous amounts of information and an expert to operate on it. To develop a model of sentiment analysis, a business needs current information to construct a model. Otherwise, it is possible to use internet sources such as WordNet or public datasets namely IMDB, blogs and social media platforms. In this scenario most public datasets are not context-specific and data needs to be selected to meet the company's objectives. Additionally, owing to the absence of financing, some languages have restricted data set, it has become hard for businesses in making the language-specific models. If the businesses want to implement such system for each user, a lot of money will have to be spent due to methods that require a lot of computing power. Some models may be difficult to implement in some cases because they are too complicated or need a lot of effort to optimize them. In 2009, Netflix held a competition known as the Netflix Prize in which competitors had to create a model that would suggest user-based movies oriented on prior viewed movies, and a group called BellKor's Pragmatic Chaos earned $1 million to create a better algorithm. However, because it is too complicated and costly, Netflix did not introduce the algorithm and stopped conducting annual competition from 2009 [15].

Privacy issues also hamper the advancement in Sentiment Analysis. Businesses and researchers need to evaluate how much information they take as individuals, particularly from western countries, who value privacy without invading the daily life of consumers. Otherwise, consumers may lodge a lawsuit against businesses for the violation of their privacy, and data privacy regulations must be enacted by the government to guarantee that businesses are transparent about the use of their customers’ data. Indeed, one of the primary reasons why China can compete with the US in the future is that the individuals of China are not as worried about data privacy as the individuals of the West [16].

VI. CONCLUSION

Sentiment analysis is a prominent field based on quick computing, large volume of data & information, complicated mathematical models based on machine learning and statistics to compare customer reviews from distinct E-Commerce websites. There is a significant improvement in results these days when using social media platforms such as Twitter, Facebook and Instagram for data collection instead of data from the website which is mostly positive. Various robust machine learning algorithms are employed to predict the sentiment which usually considered as a foremost influencer for the potential and prospective customers to make effective purchase decisions. This can provide better user experience and assist businesses to make decisions or develop a model that will enhance customer relationships too. This paper researches on retail E-Commerce business as an entity but can be implemented in all aspect of industries where any comments or reviews are crucial in making the business a success or failure. Companies can assess the magnitude of product acceptance with the assistance of sentiment analysis and can develop policies to enhance their product. Individuals can also use opinion mining instruments to create purchasing choices by comparison of competing products

based not only on requirements but also the experience of the user and government views.

REFERENCES

[1] J. M. Aversa, "Spatial Big Data Analytics: The New Boundaries of Retail Location Decision-Making," *Wilfrid Laurier University* , pp. 69-77, 2019.

[2] B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lecturers on human language technologies, vol. 5, no. 1, pp. 1-167, 2012.

[3] T. Kim, "Trader sentiment on Alibaba is surging," CNBC News, London, 2016.

[4] T. Poletti, "Fitbit stock tanks as manufacturing woes, slower demand batter holiday sales," Market Watch, California, 2016.

[5] S. and Wedel, "A regression method for simultaneous fuzzy market structuring and benefit segmentation," Journal of Marketing Research, pp. 385-397, 1991.

[6] D. D, S. C and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," Fourth International Conference on Recent Trends in Computer Science & Engineering, pp. 44-49, 2016.

[7] J. Hilden, "Statistical diagnosis based on conditional independence does not require it," Computers in Biology and Medicine, vol. 14, no. 4, pp. 429-435, 1984.

[8] J. Jain, P. Panchal, N. Suryawanshi and A. P. M. A. S. Shinde, "Sentiment Analysis Using Supervised Machine Learning," Imperial Journal of Interdisciplinary Research (IJIR), vol. 2, no. 6, 2016.

[9] R. Rajput, "Review of Sentimental Analysis Methods using Lexicon Based Approach," International Journal of Computer Science and Mobile Computing, vol. 5, no. 2, pp. 159-166, 2016.

[10] R. Rajput, "Review of Sentimental Analysis Methods using Lexicon Based Approach," *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 2, pp. 159-166, 2016.

[11] R. Rajput, "Review of Sentimental Analysis Methods using Lexicon Based Approach," *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 2, pp. 159-166, 2016.

[12] L. Liilian and Pang B, "“Thumbs up? Sentiment Classification using Machine Learning Techniques” in Proceedings of EMNLP," pp. 79-86, 2002.

[13] N. Federico , A. Carlo, F. Capeci and M. Cuadros , "Sentiment Analysis on Social Media," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 919-926, 2012.

[14] E. Cambria and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, p. IEEE Intelligent Systems, 2013.

[15] L. Bing, "Sentiment Analysis: A Multi-Faceted Problem," *IEEE Intelligent Systems*, vol. 25, no. 3, pp. 76-80, 2010.

[16] R. Wahome, "This Is How Twitter Sees The World : Sentiment Analysis Part One," 2018. [Online]. Available: https://towardsdatascience.com/the-real-world-as-seen-on-twitter-sentiment-analysis-part-one-5ac2d06b63fb. [Accessed 25 february 2019].

[17] R. Holiday, "What the Failed $1M Netflix Prize Says About Business Advice," Forbes, New Jersey, 2012.

[18] H. Waguih, "A Data Mining Approach for the Detection of Denial of Service Attack," *IAES International Journal of Artificial Intelligence*, vol. 2, no. 2, pp. 99-106, 2013.